

## **BAB II**

### **LANDASAN TEORI**

#### **2.1. Udara**

Udara adalah suatu campuran gas yang terdapat pada lapisan atmosfer yang mengelilingi bumi. Udara juga merupakan atmosfer yang berada di sekeliling bumi yang memiliki fungsi sangat penting bagi kelangsungan hidup makhluk hidup di muka bumi ini baik itu tumbuhan maupun hewan. Oleh karena itu, udara menjadi penyokong utama bagi kehidupan yang berlangsung di bumi kita ini.

Udara yang ada di bumi ini mengandung 78% nitrogen, 21% oksigen, dan 1% uap air, karbon dioksida, dan gas-gas lain (Wikipedia, 2013). Dalam udara yang ada di bumi juga terdapat bermacam-macam zat yang terkandung, salah satu contoh zat tersebut adalah oksigen ( $O_2$ ) yang digunakan untuk bernafas bagi manusia dan hewan, karbondioksida ( $CO_2$ ) digunakan untuk proses fotosintesis bagi tumbuhan dan ozon berfungsi untuk menahan radiasi pancaran sinar ultraviolet yang dikeluarkan oleh sinar matahari.

##### **2.1.1. Pencemaran Udara**

Berdasarkan Undang-Undang Nomor 23 tahun 1997 pasal 1 ayat 12 mengenai Pencemaran Lingkungan, yang mana Pencemaran udara adalah pencemaran yang disebabkan oleh aktivitas manusia seperti pencemaran yang berasal dari pabrik, kendaraan bermotor, pembakaran sampah, sisa pertanian, dan peristiwa alam seperti kebakaran hutan, letusan gunung api yang mengeluarkan debu, gas, dan awan panas. Pencemaran udara lingkungan yang dilakukan oleh aktivitas manusia dalam jumlah yang besar akan memberikan dampak yang berpengaruh bagi keseimbangan kelangsungan hidup di bumi.

Selain itu menurut Peraturan Pemerintah RI No. 41 Tahun 1999 tentang pencemaran udara yaitu peristiwa masuknya atau dimasukkannya zat, energi, dan komponen lain ke dalam udara ambien oleh kegiatan manusia, sehingga mutu udara ambien turun sampai ke tingkat tertentu yang

menyebabkan udara ambien tidak dapat memenuhi fungsinya. Jadi, pencemaran udara yang banyak terjadi diakibatkan oleh aktifitas manusia itu sendiri yang membuat mutu dari kualitas udara ambient turun pada tingkatan tertentu.

Prinsip dari pencemaran udara yang terjadi ini adalah apabila dalam udara terdapat suatu unsur-unsur pencemar (polutan) yang bersumber dari aktifitas alam dan juga kebanyakan dari aktifitas manusia. Pencemaran udara tersebut dapat mempengaruhi keseimbangan kualitas udara normal dan mengakibatkan gangguan terhadap kelangsungan hidup makhluk hidup yang ada di muka bumi.

### **2.1.2. Parameter Pencemaran Udara**

Adapun jenis parameter pencemaran udara, yang mana meliputi (Laboratorium Udara Pekanbaru, 2013) :

a) **Nitrogen Dioksida ( $\text{NO}_2$ )**

Nitrogen dioksida ( $\text{NO}_2$ ) merupakan gas yang toksik atau bersifat racun bagi manusia. Pada umumnya gas ini dapat menimbulkan gangguan pada sistem pernapasan manusia.  $\text{NO}_2$  atau nitrogen dioksida ini dapat masuk ke paru-paru dan membentuk senyawa asam nitrit ( $\text{HNO}_2$ ) dan Asam Nitrat ( $\text{HNO}_3$ ) yang merusak jaringan mukosa pada tubuh.

b) **Sulfur Dioksida ( $\text{SO}_2$ )**

Gas  $\text{SO}_2$  yang ada di udara dapat menyebabkan iritasi saluran pernapasan dan kenaikan sekresi mukosa. Konsentrasi  $\text{SO}_2$  dengan tingkat sebesar 500 ppm dapat menyebabkan kematian pada manusia. Pencemaran  $\text{SO}_2$  yang cukup tinggi akan menimbulkan malapetaka yang cukup serius hingga kematian.

c) **Carbon Monoksida ( $\text{CO}$ )**

Keracunan yang disebabkan oleh gas karbon monoksida ( $\text{CO}$ ) dapat ditandai dari keadaan ringan berupa pusing, sakit kepala, dan mual. Keracunan yang berada pada level tingkat tinggi dapat mrnyebabkan menurunnya kemampuan gerak pada tubuh, gangguan pada sistem *cardiovaskuler*, serangan jantung hingga kematian.

d) **Partikulat Matter (PM10)**

Partikel udara dalam wujud padat yang berdiameter kurang dari 10 µm yang biasanya disebut dengan PM10 (*particulate matter*) atau dikenal dengan partikel debu. Menurut para pakar lingkungan dan kesehatan masyarakat terjadinya infeksi saluran pernafasan, dikarenakan partikel padat PM10 dapat mengendap pada saluran pernafasan daerah *bronki* dan *alveoli* yang ada dalam paru-paru. PM10 sangat memprihatinkan karena memiliki kemampuan yang lebih besar untuk menembus masuk ke dalam paru-paru manusia.

e) **Ozon (O<sub>3</sub>)**

Ozon pada saat ini menjadi suatu isu aktual yang mana berkaitan dengan efek global pencemaran udara yaitu penipisan lapisan Ozon di atmosfer atas bumi kita. Ozon merupakan salah satu pencemar udara yang terus meningkat konsentrasinya. Ozon pada konsentrasi 0,3 ppm dapat menyebabkan iritasi terhadap hidung dan tenggorokan. Kontak dengan ozon pada konsentrasi 1,0–3,0 ppm selama 2 jam dapat mengakibatkan pusing berat dan kehilangan koordinasi pada beberapa orang yang sensitif. Sedangkan kontak dengan konsentrasi 9,0 ppm selama beberapa waktu dapat mengakibatkan *endema pulmonari* pada kebanyakan orang.

### 2.1.3. Indeks Kualitas Udara

Indeks Kualitas Udara digunakan sebagai bahan informasi bagi masyarakat tentang kualitas udara ambient di lokasi dan waktu tertentu. ISPU juga digunakan sebagai bahan pertimbangan pemerintah pusat dan pemerintah daerah dalam melaksanakan pengelolaan dan pengendalian pencemaran udara yang terjadi (Laboratorium Udara Pekanbaru, 2013).

Perhitungan nilai serta penomoran ISPU itu sendiri menggunakan metode perhitungan yang telah ditetapkan melalui Keputusan Kepala Badan Pengendalian Dampak Lingkungan No. KEP-107/KABAPEDAL/11/1997, yaitu :

$$I = \frac{Ia - Ib}{Xa - Xb} (Xx - Xb) + Ib \quad (2.1)$$

Dimana ;

I = ISPU

Ia = ISPU batas atas

Ib = ISPU batas bawah

Xa = Ambien batas atas

Xb = Ambien batas bawah

Xx = Ambien hasil pengukuran

Untuk nilai batas atas dan batas bawah nilai ISPU dapat dilihat pada tabel 2.1 dibawah ini.

**Tabel 2.1 Batas ISPU**

ISPU			PM10 µg/m <sup>3</sup>		SO <sub>2</sub> µg/m <sup>3</sup>		CO µg/m <sup>3</sup>		O <sub>3</sub> µg/m <sup>3</sup>		NO <sub>2</sub> µg/m <sup>3</sup>	
Kategori	Ib	Ia	Xb	Xa	Xb	Xa	Xb	Xa	Xb	Xa	Xb	Xa
50	1	50	0	50	0	80	0	5	0	5	0	0
100	51	100	51	150	81	365	6	10	121	235	0	0
200	101	199	151	350	366	800	11	17	236	400	0	1130
300	200	299	351	420	801	1600	18	34	401	800	1131	2260
400	300	399	421	500	1601	2100	35	46	801	1000	2261	3000
500	400	500	501	600	2101	2620	47	57,5	1001	1200	3001	3750

(Sumber : Peraturan Pemerintah No. 41 Tahun 1999)

Nilai ISPU dihitung untuk semua parameter yang terukur, sehingga diperoleh Nilai ISPU untuk Masing-masing parameter kualitas udara. Nilai yang diambil sebagai nilai akhir Indeks Standar Pencemaran Udara (ISPU) untuk pengukuran kualitas udara yang terjadi adalah salah satu nilai ISPU yang tertinggi dari hasil perhitungan terhadap semua parameter kualitas udara (Laboratorium Udara Pekanbaru, 2013).

Tabel Angka dan Kategori Nilai ISPU dapat dilihat pada tabel 2.2 berikut :

**Tabel 2.2 Angka dan Kategori ISPU**

INDEKS	KATEGORI
1 – 50	B a i k
51 – 100	Sedang
101 – 199	Tidak Sehat
200 – 299	Sangat Tidak Sehat
300 – lebih	Berbahaya

Skala warna tingkat kualitas udara berdasarkan Keputusan Kepala Badan Pengendalian Dampak Lingkungan No. KEP-107/KABAPEDAL/11/1997.

Scale due to norm: KEP.107/KABAPEDAL/11/1997				
0 - 50	51 - 100	101 - 199	200 - 299	300 - 500
BAIK	SEDANG	TIDAK SEHAT	SANGAT TIDAK SEHAT	BERBAHAYA
GOOD	MODERATE	UNHEALTHY	VERY UNHEALTHY	DANGEROUS

**Gambar 2.1 Skala Warna Kualitas Udara**

Keterangan :

- Baik** : Tingkat kualitas udara yang tidak memberikan efek bagi kesehatan manusia atau hewan dan tidak berpengaruh pada tumbuhan, bangunan atau nilai estetika.
- Sedang** : Tingkat kualitas udara yang tidak berpengaruh pada kesehatan manusia ataupun hewan tetapi berpengaruh pada tumbuhan yang sensitif dan nilai estetika.
- Tidak Sehat** : Tingkat kualitas udara yang bersifat merugikan pada manusia ataupun kelompok hewan yang sensitif atau bisa menimbulkan kerusakan pada tumbuhan ataupun nilai estetika.
- Sangat Tidak Sehat** : Tingkat kualitas udara yang dapat merugikan kesehatan pada sejumlah segmen populasi yang terpapar.
- Berbahaya** : Tingkat kualitas udara berbahaya yang secara umum dapat merugikan kesehatan yang serius pada populasi.

## **2.2. Data Mining**

### **2.2.1 Definisi Data Mining**

Data *mining* merupakan suatu proses menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* (mesin pembelajaran) dalam mengidentifikasi suatu informasi yang berguna dan pengetahuan (*Knowledge*) berasal dari berbagai database yang besar (Turban, 2005).

Ada beberapa definisi tentang *data mining* menurut (Kusnawi, 2007), yaitu antara lain :

1. *Data mining* adalah serangkaian suatu proses yang bertujuan untuk mengali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.
2. *Data mining* adalah analisa otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan suatu pola atau kecendrungan yang penting yang tidak disadari keberadaannya
3. *Data mining* atau *Knowledge Discovery in Database* (KDD) adalah pengalian informasi yang tersembunyi, dimana pola informasi tersebut sebelumnya tidak dikenali dan memiliki potensi yang berguna sebagai informasi. *Knowledge Discovery in Database* (KDD) merupakan proses menentukan yang berguna serta menemukan pola-pola yang ada dalam data. *Data mining* adalah salah satu langkah dari serangkaian proses iterasi dalam *Knowledge Discovery in Database* (KDD).

### **2.2.2 Proses Data Mining**

Dalam menemukan suatu model dari data dalam *data mining* ada beberapa proses atau prosedur yang harus dilakukan, yaitu :

1. Merumuskan permasalahan dalam proses atau prosedur ini yang ditetapkan dalam rumusan masalah dan variable-variabel yang terlibat.
2. Pengumpulan data yang berkonsentrasi pada proses pembuatan data dan pengumpulan data.
3. *Pra Process data* bertujuan untuk menyeleksi jenis data-data apa saja yang digunakan dalam proses data mining. Adapun yang dilakukan pada tahapan *pra process data* ini antara lain :

a. *Selection data*

Proses ini dilakukan untuk pemilihan kumpulan data, dan menciptakan target data atau memfokuskan apa yang akan dijadikan data sampel.

b. *Cleaning data*

*Cleaning data* dilakukan membuang data yang tidak konsisten dan *noise*.

c. *Transformasi data*

Proses ini mentransformasikan atau menggabungkan data ke dalam yang lebih tepat untuk melakukan proses *mining*.

4. Mengestimasi model seleksi dan implementasi terhadap metode *data mining* yang tepat dalam proses utama dengan menggunakan teknik, metode, algoritma tertentu.
5. Menafsirkan pola-pola yang dihasilkan, mengevaluasi apakah pola tersebut ditemukan atau tidak dan kemudian menarik kesimpulan.

### 2.2.3 Teknik Data Mining

Teknik *Data mining* ini memiliki beberapa teknik yang berdasarkan tugas yang dimilikinya:

1. *Description* (Deskripsi)

Deskripsi ini dipakai untuk menemukan cara mendeskripsikan pola dan trend yang tersembunyi dalam data.

2. *Estimation* (Estimasi)

Estimasi ini hampir mirip dengan klasifikasi, kecuali variable tujuan yang lebih ke arah numerik dari pada pengkategorian.

3. *Prediction* (Prediksi)

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja prediksi hasilnya menunjukkan suatu yang belum terjadi (mungkin belum terjadi di masa depan).

4. *Classification* (Klasifikasi)

Klasifikasi merupakan salah satu teknik *data mining* yang melakukan klasifikasi data variable dengan tujuan yang bersifat pengkategorian.

5. *Clustering* (Pengelompokan)

*Clustering* ini lebih ke arah pengelompokkan record, pengamatan, atau kasus dalam kelas yang memiliki kemiripan.

6. *Asotiation* (Asosiasi)

Asosiasi bertujuan mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu dengan menekankan sebuah kelas masalah yang dicirikan dengan analisi keranjang pasar.

### 2.3. Teknik Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan “jika-maka”, berupa *decision tree*, *formula matematis* atau *neural network*. Teknik klasifikasi juga bertujuan untuk melakukan pemetaan data ke dalam kelas yang sudah didefinisikan sebelumnya berdasarkan nilai atribut yang dimiliki data tersebut (Hans & Kamber, 2006).

Klasifikasi merupakan proses menyatakan suatu objek ke dalam salah satu kategori yang sudah didefinisikan sebelumnya. Tahapan-tahapan klasifikasi sebagai berikut :

1) Pembangunan Model

Dalam tahap ini bertujuan untuk menyelesaikan masalah klasifikasi data, model ini dibangun berdasarkan *training set*.

2) Penerapan Model

Penerapan model yang sudah dibangun sebelumnya akan digunakan untuk menentukan atribut atau kelas dari sebuah data baru.

3) Evaluasi

Dalam tahap ini melakukan evaluasi menggunakan parameter terukur untuk menentukan apakah model tersebut diterima.



Ada beberapa metode atau algoritma klasifikasi dalam data mining, antara lain (Han & Kamber, 2006) :

### **1. Algoritma C5.0**

Merupakan algoritma klasifikasi data mining yang mana pemilihan atribut yang akan diproses menggunakan *information gain* sebagai kriterianya. Algoritma ini memiliki fitur penting yang membuat algoritma C5.0 ini lebih unggul dari algoritma terdahulunya, dan fitur tersebut antara lain :

- a. Algoritma C5.0 dirancang untuk menganalisis basis data substansial yang berisi puluhan sampai ratusan *record* dan satuan hingga ratusan *field* numeric dan nominal.
- b. Memaksimumkan tingkat penafsiran pengguna terhadap hasil yang disajikan dalam dua bentuk yaitu menggunakan pohon keputusan (*decision tree*) dan sekumpulan aturan *IF-Then* yang mudah dimengerti.
- c. Algoritma C5.0 mudah digunakan dan tidak membutuhkan pengetahuan yang tinggi tentang statistik atau *machine learning*.

### **2. Algoritma Decision Tree**

Algoritma *decision tree* adalah algoritma yang menggunakan konsep diagram alir yang mirip dengan struktur pohon, dimana internal node menotasikan atribut yang diuji, setiap cabang nya merepresentasikan hasil dari atribut tes tersebut, dan *leaf node* merepresentasikan kelas-kelas tertentu atau distribusi kelas-kelas. Pada umumnya ada beberapa cirri kasus yang cocok untuk diterapkan dengan algoritma ini, antara lain:

- a. Data dinyatakan dengan pasangan atribut dan nilainya Misalnya atribut temperature dan nilainya adalah dingin.
- b. Label/output data biasanya bernilai *diskrit*. Outputnya bias bernilai ya atau tidak, sakit atau tidak sakit, diterima atau ditolak.
- c. Data mempunyai *missing value*. Misalkan untuk beberapa data, nilai dari suatu atribut nya tidak diketahui. Maka dalam keadaan ini *decision tree* masih mampu member solusi yang baik.

### 3. Algoritma *K-Nearest Neighbor* (K-NN)

Algoritma *K-Nearest Neighbor* adalah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya terdekat dengan objek tersebut. Data pembelajarannya diproyeksikan ke ruang berdimensi banyak, dimana merepresentasikan fitur dari data.

## 2.4 Algoritma K-NN(*k-Nearest Neighbor*)

### 2.4.1 Definisi Algoritma *K-Nearest Neighbor*

Algoritma K-NN adalah suatu metode yang menggunakan algoritma *supervised* (Nugroho, 2011). Perbedaan antara *supervised learning* dengan *unsupervised learning* yaitu pada *supervised learning* bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Sedangkan *unsupervised learning*, data belum memiliki pola apapun, dan tujuan *unsupervised learning* untuk menemukan pola dalam data.

Tujuan dari algoritma K-NN adalah untuk mengklasifikasi objek baru berdasarkan atribut dan *training samples* (Larose D, 2005). Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN.

### 2.4.2 Proses *K-Nearest Neighbor*

Prinsip kerja *K-Nearest Neighbor* adalah mencari jarak antara dua titik yaitu titik *training* dan titik *testing*, yang kemudian dilakukan evaluasi dengan k tetangga terdekatnya dalam data *training* (pelatihan). Persamaan perhitungan untuk mencari jarak dengan menggunakan rumus *euclidean* :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.2)$$

Dimana, d adalah jarak antara titik pada data training x dan titik data testing y yang akan diklasifikasikan, dimana  $x = x_1, x_2, \dots, x_i$  dan  $y = y_1, y_2, \dots, y_i$  dan  $I$  merepresentasikan nilai atribut serta  $n$  merupakan dimensi data atribut (Hans & Kamber, 2006).

## 2.5 Penerapan Algoritma *K-Nearest Neighbor*

Sebelum melakukan penerapan algoritma *k-nearest neighbor*, ada hal yang harus diperhatikan terlebih dahulu yaitu data *training* (sampel) dan data *testing* (uji) sudah terlebih dahulu ditentukan sebelum dilakukannya proses perhitungan dengan *euclidean distance*. Kemudian baru dilakukan tahapan atau langkah dalam melakukan penerapan algoritma *K-Nearest Neighbor* seperti gambar *flowchart 2.2* :



**Gambar 2.2 Flowchart Algoritma K-NN**

## 2.6 Pengujian Algoritma *K-NN (k-Nearest Neighbor)*

Pengujian kinerja sistem klasifikasi pada algoritma *k-nearest neighbor* ini dapat dilakukan dengan menggunakan *confusion matrix*. *Confusion matrix* ini alat yang berguna untuk menganalisis seberapa baiknya *classifier* yang kita pakai dapat mengenali pola dari kelas yang berbeda (Hans & Kamber, 2006). Pengukuran dalam menentukan klasifikasi ini dapat dilakukan dengan :

1. *Accuracy* (akurasi)  
Mengukur tingkat keberhasilan berdasarkan kedekatan antara nilai prediksi dengan nilai aktualnya.
2. *Error*  
Mengukur rata-rata kegagalan dalam melakukan klasifikasi antara nilai prediksi dan nilai actual.

Secara umum, *accuracy* dan *error rate* dapat dirumuskan sebagai berikut :

**Tabel 2.2 Model Penentuan Keputusan (*Confusion Matrix*)**

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai	TRUE	TP ( <i>True Positive</i> ) <i>Correct result</i>	FP ( <i>False Positive</i> ) <i>Unexpected result</i>
	FALSE	FN ( <i>False Negative</i> ) <i>Missing result</i>	TN ( <i>True Negative</i> ) <i>Correct absence of result</i>

Keterangan ;

**1. TP (*True Positive*)**

Apabila kenyataan dan sistem menghasilkan hasil yang positif.

**2. TN (*True Negative*)**

Apabila kenyataan dan sistem menghasilkan hasil yang negative.

**3. FP (*False Positive*)**

Apabila kenyataan negatif, tetapi sistem memutuskan positif.

**4. FN (*False Negative*)**

Apabila kenyataan positif, tetapi sistem memutuskan negative.

Rumus persamaan *accuracy* (akurasi) dan *error rate* dapat dilihat dibawah ini :

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \% \quad (2.3)$$

$$\text{error} = \frac{FP + FN}{TP + TN + FP + FN} \times 100 \% \quad (2.4)$$